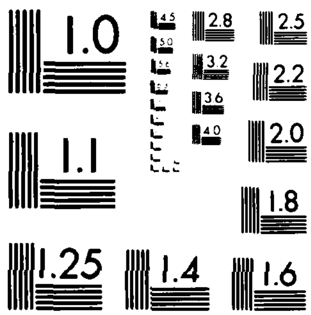

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

(13)

AD-A145 476

CONSTRAINED OPTIMIZATION
USING ITERATED PARTIAL KUHN-TUCKER VECTORS

by

Richard L. Dykstra

and

Peter C. Wollan

Department of Statistics and Actuarial Science
The University of Iowa
Iowa City, IA 52242

Technical Report #105
May 1984

This research was partially supported by ONR Contract N00014-83-K-0249.

ABSTRACT

A frequently occurring problem is that of minimizing a convex function subject to a finite set of inequality constraints. Often what makes this problem difficult is the sheer number of constraints. That is, we could solve this problem for a smaller set of constraints, but solving for the total set causes difficulty. Here we discuss an approach which uses our ability to solve these partial problems to lead to a total solution. We will illustrate the method with several examples in the last section of the paper. ~~Our~~ approach will be somewhat heuristic in nature to promote understanding.

Key Words and Phrases: Optimization, Kuhn-Tucker vectors, inequality constraints, iteration, least squares, I-projections.

1. INTRODUCTION.

Let us consider the following convex programming problem: Minimize the convex function $f_0(x)$ (defined on some subset of R^n) over the region C subject to the constraints $f_1(x) \le 0, \cdots, f_m(x) \le 0$ where the f_i, $i = 1, \cdots, m$ are finite convex functions on C. We could of course define $f_0(x)$ to be $+\infty$ for $x \notin C$, and hence assume that our functions are defined over R^n. We shall call $f_0(x)$ the objective function, and refer to $f_i(x)$, $i = 1, \cdots, m$ as the constraint functions.

We shall define $\lambda = (\lambda_1, \cdots, \lambda_m) \in R^m$ to be a vector of Kuhn-Tucker coefficients for our problem, or simply a Kuhn-Tucker vector if $\lambda_i \ge 0$, $i = 1, \cdots, m$ and if

$$\inf_{x \in C} f_0(x) + \lambda_1 f_1(x) + \cdots + \lambda_m f_m(x)$$

is finite and equal to the optimal value of our original problem. The existence of Kuhn-Tucker (KT) vectors is guaranteed under mild conditions. (See Rockafellar (1970) for a discussion of this material.) Part of the importance of KT vectors stems from the fact that they can convert a constrained problem into an unconstrained (or at least more simply constrained) problem.

Another important construct is the Lagrangian associated with our problem. It is defined as the function L on

$R^m \times R^n$ given by

$$L(\lambda,x) = \begin{cases} f_0(x) + \lambda_1 f_1(x) + \cdots + \lambda_m f_m(x), & x \in C,\ \lambda_i \geq 0,\ i = 1,\cdots,m \\ -\infty, & x \in C,\ \lambda_i < 0\ (\exists i) \\ +\infty, & x \notin C. \end{cases}$$

A vector pair (λ^0, x^0) is said to be a saddle-point of L if

$$(1.1) \qquad L(\lambda, x^0) \leq L(\lambda^0, x^0) \leq L(\lambda^0, x) \quad \forall x, \lambda.$$

Such saddle-points play a big role in the following fundamental theorem (Rockafellar, p. 281).

<u>Theorem</u> 1.1. In order that λ^0 be a KT vector and x^0 be an optimal solution, it is necessary and sufficient that (λ^0, x^0) be a saddle-point of the Lagrangian L. Moreover, this condition holds iff x^0 and λ^0 satisfy

$$(1.2) \quad \begin{array}{ll} \text{(a)} & \lambda_i^0 \geq 0,\ f_i(x^0) \leq 0 \ \text{ and } \ \lambda_i^0 f_i(x^0) = 0,\ i = 1,\cdots,m \\ \text{(b)} & 0 \in [\partial f_0(x^0) + \lambda_1 \partial f_1(x^0) + \cdots + \lambda_m \partial f_m(x^0)]. \end{array}$$

The notation $\partial f(x^0)$ indicates the set of subgradients of f at x_0. Condition (b) implies that x^0 minimizes $f_0(x) + \lambda_1 f_1(x) + \cdots + \lambda_m f_m(x)$. Of course if the f_i are all differentiable, (1.2)(b) may be replaced by

(1.2) (b′) $\nabla f_0(x^0) + \lambda_1 \nabla f_1(x^0) + \cdots + \lambda_m \nabla f_m(x^0) = 0.$

Given that one can solve a constrained minimization problem, condition (1.2)(b′) can often be used to find a KT vector for the problem. This fact shall prove useful later on.

2. THE METHOD.

The basic idea behind our approach is that one can reduce the number of constraints being considered by modifying the objective function through the use of estimated KT vectors. At each stage, the modified problem is solved, and updated estimates of KT vectors are found. Under fairly general conditions, the solutions to the modified problems must converge to a true global solution.

To be more specific, we assume that our constraint functions are grouped into vectors and given by

$$\begin{aligned} f_1(x) &= (f_{11}(x), f_{12}(x), \cdots, f_{1m_1}(x)) \\ &\vdots \\ f_k(x) &= (f_{k1}(x), f_{k2}(x), \cdots, f_{km_k}(x)) \end{aligned}$$

where $\sum_{i=1}^{k} m_i = m$.

We then define our Lagrangian as

$$L(\boldsymbol{\lambda},x) = \begin{cases} f_0(x) + \boldsymbol{\lambda}_1' f_1(x) + \cdots + \boldsymbol{\lambda}_k' f_k(x), & x \in C,\ \lambda_{ij} \geq 0,\ \forall ij \\ -\infty, & x \in C,\ \lambda_{ij} < 0,\ \exists i,j \\ +\infty, & x \notin C, \end{cases}$$

where $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1,\cdots,\boldsymbol{\lambda}_k)$ and $\boldsymbol{\lambda}_i$ is an $m_i \times 1$ vector.

We shall use $L_i(\boldsymbol{\lambda},x)$ to denote $L(\boldsymbol{\lambda},x)$ with $\boldsymbol{\lambda}_i$ set equal to zero, and $L_i(\boldsymbol{\lambda}_i,x|\boldsymbol{\lambda})$ to denote $L(\boldsymbol{\lambda},x)$ considered as a function of $\boldsymbol{\lambda}_i$ and x $(\boldsymbol{\lambda}_1,\cdots,\boldsymbol{\lambda}_{i-1},\boldsymbol{\lambda}_{i+1},\cdots,\boldsymbol{\lambda}_k$ are regarded as fixed).

Initially, we set $\boldsymbol{\lambda}_j^{(0,j)} = 0,\ j = 1,\cdots,k$ and $\boldsymbol{\lambda}^{(1,0)} = (\boldsymbol{\lambda}_1^{(0,1)},\cdots,\boldsymbol{\lambda}_k^{(0,k)})$. Our algorithm is sequentially defined as follows (beginning with $n = 1,\ i = 0$):

a) Let $x^{(n,i+1)}$ $(x^{(n+1,1)}$ if $i = k)$ denote the solution to

$$\begin{gathered} \text{Minimize } L_{i+1}(\lambda^{(n,i)},x) \\ x: f_{i+1,j}(x) \leq 0\ \forall j \\ (x: f_{1j}(x) \leq 0\ \forall j \text{ if } i = k) \end{gathered}$$

This is a **convex** programming problem which we assume we can solve. We let $\boldsymbol{\lambda}_{i+1}^{(n,i+1)}$ $(\boldsymbol{\lambda}_1^{(n+1,1)}$ if $i = k)$ denote a KT vector for the problem. (Condition (1.2) may prove useful in obtaining this KT vector.)

b) We now update our global KT estimate by setting

$$\lambda^{(n,i+1)} = (\lambda_1^{(n,1)}, \ldots, \lambda_{i+1}^{(n,i+1)}, \lambda_{i+2}^{(n-1,i+2)}, \ldots, \lambda_k^{(n-1,k)}).$$

(If $i = k$, we set

$$\lambda^{(n+1,1)} = (\lambda_1^{(n+1,1)}, \lambda_2^{(n,2)}, \ldots, \lambda_k^{(n,k)}.)$$

We then replace (n,i) by $(n,i+1)$ ($(n+1,1)$ if $i = k$) and return to step a).

In many situations, the problems in step a) which must be solved are always of the same form, and hence lead to easily written computer programs for performing these steps. We shall give some explicit examples in Section 4.

3. JUSTIFICATION FOR THE ALGORITHM.

The crucial fact that justifies this procedure is that the Lagrangian can only increase at each step in the algorithm. This follows since (for $i < k$)

$$\begin{aligned} L(\lambda^{(n,i)}, x^{(n,i)}) &= L_i(\lambda_i^{(n,i)}, x^{(n,i)} | \lambda^{(n,i)}) \\ &\le L_i(\lambda_i^{(n,i)}, x^{(n,i+1)} | \lambda^{(n,i)}) \text{ (by Thm. 1.1)} \\ &= L_{i+1}(\lambda_{i+1}^{(n,i)}, x^{(n,i+1)} | \lambda^{(n,i+1)}) \\ &\le L_{i+1}(\lambda_{i+1}^{(n,i+1)}, x^{(n,i+1)} | \lambda^{(n,i+1)}) \text{ (by Thm. 1.1)} \\ &= L(\lambda^{(n,i+1)}, x^{(n,i+1)}). \end{aligned}$$

A similar argument holds if $i = k$ for showing

$$L(\lambda^{(n,k)}, x^{(n,k)}) \le L(\lambda^{(n+1,1)}, x^{(n+1,1)}).$$

Moreover, we note that if $y \in C$ is any vector such that $f_{ij}(y) \le 0, \ \forall i,j,$ then

$$(3.1) \qquad L(\lambda^{(n,i)}, x^{(n,i)}) \le L(\lambda^{(n,i)}, y) \le f_0(y),$$

so that $\lim L(\lambda^{(n,i)}, x^{(n,i)})$ exists finite independently of i. Now, since $x^{(n,i)}$ minimizes a convex function, under conditions which guarantee sufficient curvature of $L(x,\lambda)$ (such as $xHx \ge \gamma\|x\|^2$ for some $\gamma > 0$ where H is the Hessian of f_0), we know that $x^{(n,i+1)}$ must be close to $x^{(n,i)}$ for sufficiently large n. Now, if $x^{(n,i)}$ must contain a convergent subsequence (such as if C is a bounded region) converging to $x^0 \in C$, then we only need continuity properties of f_0 and f_i, $i = 1,\cdots,m$ to guarantee that $f_i(x^0) \le 0$, $i = 1,\cdots,m$ and $f_0(x^0) \le f_0(y)$ for all y in C which satisfy all constraints (by 3.1). Thus x^0 must be a solution to our problem, and since every convergent subsequence converges to x^0, the algorithm must work correctly.

4. APPLICATIONS.

1. Let us first consider least squares problems under linear inequality constraints. Thus, we wish to minimize the objective function

$$f_0(x) = \frac{1}{2}\sum_{1}^{n}(g_i - x_i)^2 w_i$$

subject to the constraints

$$f_i(x) = x'a_i = \sum_{j=1}^{n} a_{ij}x_j \le 0,\ i = 1,\cdots,m,$$

where $w > 0$, $g, a_1, \cdots, a_m$ are given $n \times 1$ vectors such that there exists at least one vector x where $a_i'x \le 0\ \forall i$.

Of course, the solution to our problem under a single constraint $a_i'x \le 0$ can be easily found by the expression

$$(4.1)\qquad P_i(g) = \begin{cases} g & , \text{ if } \sum_{j=1}^{n} a_{ij}g_j \le 0 \\ (g_1', \cdots, g_n'), & \text{ if } \sum_{j=1}^{n} a_{ij}g_j > 0, \end{cases}$$

where

$$g_j' = g_j - \left(\sum_{\ell=1}^{n} g_\ell a_{i\ell}\right)a_{ij}w_j^{-1} \Big/ \sum_{\ell=1}^{n} a_{i\ell}^2 w_\ell^{-1}.$$

The corresponding KT value can then be found from (1.2)(b') as

$$\lambda_i = (g_j - P_i(g)_j)w_j a_{ij}^{-1}, \tag{4.2}$$

where j is any index such that $a_{ij} \neq 0$. Suppose now that we modify our objective function (with g replaced by an arbitrary h) by adding $\lambda_i f_i(x)$. However, the problem

$$\underset{x' a_r \leq 0}{\text{Minimize}} \ \frac{1}{2} \sum_{j=1}^{n} (h_j - x_j)^2 w_j + \lambda_i f_i(x) \tag{4.3}$$

has the same solution as

$$\underset{x' a_r \leq 0}{\text{Minimize}} \ - \sum_{j=1}^{n} x_j w_j [h_j + (P_i(g)_j - g_j)] + \frac{1}{2} \sum_{j=1}^{n} x_i^2 w_i$$

which is equivalent to

$$\underset{x' a_r \leq 0}{\text{Minimize}} \ \frac{1}{2} \sum_{j=1}^{n} (h_j + (P_i(g)_j - g_j) - x_j)^2 w_j . \tag{4.4}$$

Thus our adjustment still leaves the problem as the same type of least squares problem, but with the value h modified to now be $h + (P_i(g) - g)$.

We can now apply the method proposed in section 2, which can be expressed as follows:

1) Set $g_{11} = P_1(g)$, and $I_{11} = g_{11} - g$.
2) Set $g_{12} = P_2(g_{11})$, and $I_{12} = g_{12} - g_{11}$.
3) Continue, until $g_{1m} = P_m(g_{1,m-1})$, and $I_{1m} = g_{1m} - g_{1,m-1}$.

4) Now set $g_{21} = P_1(g_{1m}-I_{11})$, and $I_{21} = g_{21}-(g_{1m}-I_{11})$.

5) Contine. In general, set $g_{nj} = P_j(g_{n,j-1}-I_{n-1,j})$, and $I_{nj} = g_{nj}-(g_{n,j-1}-I_{n-1,j})$ if $j > 1$, and $g_{n1} = P_1(g_{n-1,m}-I_{n-1,1})$ and $I_{n1} = g_{n1}-(g_{n-1,m}-I_{n-1,1})$.

This scheme is easy to program since the projections are of such simple form, and there is no branching or searching involved. Moreover, $g_{n,j}$ is guaranteed to converge to the true solution. This is a special case of an algorithm given by Dykstra (1983).

2. Depending upon the nature of the linear constraints, it may be possible to easily find the projection under several simultaneous constraints. For example, if

$$f_i(x) = x_i - x_{i+1}, \quad i = 1,\cdots,n-1,$$

the set of vectors which satisfy all $n-1$ constraints are just the nondecreasing vectors. Projections onto these types of regions are quite tractable (see Barlow, Bartholomew, Bremner and Brunk (1972)). Another set of linear inequality constraints which can be simultaneously handled are

$$f_i(x) = i^{-1}\sum_{j=1}^{i} x_j - x_{i+1}, \quad i = 1,\cdots,n-1$$

(Shaked (1979) and Dykstra and Robertson (1983)).

By being able to handle large constraint sets, we can improve the efficiency of our method.

To elaborate, we consider the problem

$$(4.5) \qquad \underset{x'A_i\le 0,\, i=1,\cdots,k}{\text{Minimize}} \ \tfrac{1}{2}\sum_{1}^{n}(g_j - x_j)^2 w_j$$

where $A_i = (a_{i1},\cdots,a_{im_i})$ is an $n \times m_i$ matrix of independent columns. We assume that we can solve (4.5) for any particular i and any g, and will denote the solution by $P_i(g)$. It can be shown that the KT vector for this problem is

$$\lambda_i^{m_i \times 1} = (A_i' A_i)^{-1} A_i' (g\cdot w - P_i(g)\cdot w)$$

where $x\cdot y$ denotes coordinatewise multiplication.

Interestingly enough, we can repeat the argument used in (4.3) and (4.4) to derive the same type of result. Thus it follows that our earlier stated algorithm is still valid if $P_i(g)$ denotes the projection of g onto $\{x; x'A_i \le 0\}$.

This extension may prove very useful for situations where there are a great many constraints. For example, Dykstra and Robertson (1982) were able to find least squares projections of rectangular arrays under the constraints of nondecreasing rows and nondecreasing columns for even large arrays.

3. As a final example, we consider the problem of finding I-projections onto an intersection of linear inequality regions.

We will not elaborate on the importance of I-projections, only state that they occur in a myriad of places in many different settings. We refer the reader to Kullback (1959) or Csiszar (1975) for elaboration on their importance. The problem we are dealing with is

$$(4.6) \qquad \underset{\substack{a_i' p \le c_i, i=1,\cdots,m \\ \sum_i^n p_i = 1, p_i \ge 0}}{\text{Minimize}} \quad \sum_{i=1}^{n} p_i \ln(p_i/r_i)$$

where $r \neq 0$ is a finite, nonnegative, vector and the set of feasible points is not empty. We will define our objective function as

$$f_0(p) = \begin{cases} \sum_{i=1}^{n} p_i \ln(p_i/r_i), & p_i \ge 0, \quad \sum_1^n p_i = 1 \\ \infty & , \text{ elsewhere.} \end{cases}$$

Our constraint functions are

$$f_i(p) = (a_i - c_i)' p = \sum_{j=1}^{n} (a_{ij} - c_i) p_j, \quad i = 1, \cdots, m,$$

where c_i is a vector of constants. We note that r can be arbitrarily scaled without changing the problem.

The solution to the problem of minimizing $f_0(p)$ subject

to $f_i(p) \le 0$ is given by

$$\hat{p}_{ij} = r_j \, e^{-\lambda_i(a_{ij}-c_i)} \Big/ \sum_{j=1}^{n} r_j \, e^{-\lambda_i(a_{ij}-c_i)}$$

where λ_i is the solution to

$$\sum_{j=1}^{n} (a_{ij}-c_i) r_j \, e^{-\lambda(a_{ij}-c_i)} = 0$$

providing it is nonnegative, and zero otherwise. It also turns out that λ_i is the KT value associated with the problem.

Now if we modify our objective function (with r replaced by an arbitrary nonnegative, nonzero h) by adding $\lambda_i f_i(x)$, our problem becomes

$$\underset{(a_r-c_r)'p\le 0}{\text{Minimize}} \; f_0(p) + \lambda_i \sum_{j=1}^{n} (a_{ij}-c_i) p_j ,$$

or

$$\underset{\substack{(a_r-c_r)'p\le 0 \\ p_j\ge 0 \,\forall j,\, \sum_1^n p_j=1}}{\text{Minimize}} \; \sum_{j=1}^{n} p_j [\ln p_j/h_j - \ln e^{-\lambda_i(a_{ij}-c_i)}]$$

or equivalently,

$$(4.7) \qquad \underset{\substack{(a_r-c_r)'p\le 0 \\ p_j\ge 0 \,\forall j,\, \sum_1^n p_j=1}}{\text{Minimize}} \; \sum_{j=1}^{n} p_j \ln p_j/h_j(\hat{p}_{ij}/r_j).$$

The key point is that our problem is precisely of the

same form as before, except that we have modified our vector h. Thus we may use our procedure of modifying our objective function using updated estimates of the KT vector, and only have to solve the one type of problem. Setting $\lambda_i = 0$ in (4.7) is equivalent to setting $\hat{p}_i/r \equiv 1$. This scheme is quite effective for finding I-projections under multiple linear inequality constraints.

In summary, this procedure seems to work quite well for situations where many constraints are involved, and partial solutions (solutions under partial constraints) are easily available.

REFERENCES

Barlow, R.E., Bartholomew, D.J., Bremner, J.M., and Brunk, H.D. (1972). Statistical Inference Under Order Restrictions. John Wiley, New York.

Csiszar, I. (1975). "I-Divergence Geometry for Probability Distributions and Minimization Problems," Annals of Probability, 3, 146-159.

Dykstra, Richard L. (1983). "An Algorithm for Restricted Least Squares Regression," Journal of the American Statistical Association, 78, 837-842.

Dykstra, Richard L. and Robertson, Tim (1982). "An Algorithm for Isotonic Regression for Two or More Independent Variables," Annals of Statistics, 10, 708-716.

Dykstra, Richard L. and Robertson, Tim (1983). "On Testing Monotone Tendencies," Journal of the American Statistical Association, 78, 342-350.

Kullback, S. (1959). Information Theory and Statistics. John Wiley, New York.

Rockafellar, R. Tyrrell (1970). Convex Analysis. Princeton University Press, Princeton, NJ.

Shaked, Moshe (1979). "Estimation of Starshaped Sequences of Poisson and Normal Means," Annals of Statistics, 7, 729-741.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 105	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Constrained Optimization Using Iterated Partial Kuhn-Tucker Vectors.		5. TYPE OF REPORT & PERIOD COVERED Technical - August 1984
		6. PERFORMING ORG. REPORT NUMBER 105
7. AUTHOR(s) Richard L. Dykstra Peter C. Wollan		8. CONTRACT OR GRANT NUMBER(s) N00014-83-K-0249
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics and Actuarial Science The University of Iowa Iowa City, Iowa 52242		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 N. Quincy Arlington, VA 22217		12. REPORT DATE ~~May~~ August 1984
		13. NUMBER OF PAGES 15
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Optimization, Kuhn-Tucker vectors, inequality constraints, iteration, least squares, I-projections.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A frequently occurring problem is that of minimizing a convex function subject to a finite set of inequality constraints. Often what makes ths problem difficult is the sheer number of constraints. That is, we could solve this problem for a smaller set of constraints, but solving for the total set causes difficulty. Here we discuss an approach which uses our ability to solve these partial problems to lead to a total solution. We will illustrate the method with several examples in the last section of the paper. Our approach will be somewhat heuristic in nature to promote understanding.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

4